Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution

Zahnd, C ; Sarkar, C A ; Plückthun, A

Abstract: Directed evolution is a powerful approach for isolating high-affinity binders from complex libraries. In affinity maturation experiments, binders with the highest affinities in the library are typically isolated through selections for decreased off rate using a suitable selection platform (e.g. phage display or ribosome display). In such experiments, the library is initially exposed to biotinylated antigen and the binding reaction is allowed to proceed. A large excess of unbiotinylated antigen is then added as a competitor to capture the vast majority of rapidly dissociating molecules; the slowly dissociating library members can subsequently be rescued by capturing the biotin-carrying complexes. To optimize the parameters for such affinity maturation experiments, we performed both deterministic and stochastic simulations of off-rate selection experiments using different input libraries. Our results suggest that the most critical parameters for achieving the lowest off rates after selection are the ratio of competitor antigen to selectable antigen and the selection time. Furthermore, the selection time has an optimum that depends on the experimental setup and the nature of the library. Notably, if selections are carried out for times much longer than the optimum, equilibrium is reached and the selection pressure is weakened or lost. Comparison of different selection strategies revealed that sequential selection rounds with lower stringency are favored over high-stringency selection experiments due to enhanced diversity in the selected pools. Such simulations may be helpful in optimizing affinity maturation strategies and off-rate selection experiments.

DOI: https://doi.org/10.1093/protein/gzp087

Protein Engineering, Design & Selection vol. 23 no. 4 pp. 175–184, 2010
Published online February 3, 2010 doi:10.1093/protein/gzp087

Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution

Christian Zahnd[1,4], Casim A.Sarkar[2,3,5] and Andreas Plückthun[1,5]

[1]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, Departments of [2]Bioengineering and [3]Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104-6321, USA and [4]Present address: Molecular Partners AG, Wagistrasse 14, CH-8952 Schlieren, Switzerland

[5]To whom correspondence should be addressed: casarkar@seas.upenn.edu or plueckthun@bioc.uzh.ch

Received September 9, 2009; revised November 25, 2009; accepted November 30, 2009

Edited by Andrew Bradbury

Directed evolution is a powerful approach for isolating high-affinity binders from complex libraries. In affinity maturation experiments, binders with the highest affinities in the library are typically isolated through selections for decreased off rate using a suitable selection platform (e.g. phage display or ribosome display). In such experiments, the library is initially exposed to biotinylated antigen and the binding reaction is allowed to proceed. A large excess of unbiotinylated antigen is then added as a competitor to capture the vast majority of rapidly dissociating molecules; the slowly dissociating library members can subsequently be rescued by capturing the biotin-carrying complexes. To optimize the parameters for such affinity maturation experiments, we performed both deterministic and stochastic simulations of off-rate selection experiments using different input libraries. Our results suggest that the most critical parameters for achieving the lowest off rates after selection are the ratio of competitor antigen to selectable antigen and the selection time. Furthermore, the selection time has an optimum that depends on the experimental setup and the nature of the library. Notably, if selections are carried out for times much longer than the optimum, equilibrium is reached and the selection pressure is weakened or lost. Comparison of different selection strategies revealed that sequential selection rounds with lower stringency are favored over high-stringency selection experiments due to enhanced diversity in the selected pools. Such simulations may be helpful in optimizing affinity maturation strategies and off-rate selection experiments.

Keywords: affinity maturation/kinetic simulation/off rate/off-rate selection/selection strategy

Introduction

High-throughput selection technologies have attracted significant interest over the past decade. Progress in the synthesis of large synthetic libraries of high quality (Binz *et al*., 2005; Mondon *et al*., 2008) as well as in tailoring random mutagenesis rates and strategies (Wong *et al*., 2006) has set the stage for selecting high-affinity binding molecules. Powerful selection technologies, such as phage display (Smith, 1985; Parmley and Smith, 1988), ribosome display (Hanes and Plückthun, 1997) and cell-surface display using *Escherichia coli* (Georgiou *et al*., 1997) or yeast (Boder and Wittrup, 1997), have allowed the efficient use of such large libraries. Nonetheless, experimental strategies must be carefully defined to enrich for clones with a significant improvement in the desired property, such as increased affinity. The stringency of such a selection pressure is crucial for the outcome of any selection approach. The selection for binders with low-to-mid nanomolar affinity can usually be accomplished by 'panning' experiments in which binders are enriched from a library by exposing it to surface-immobilized antigen (Hanes *et al*., 1998; Jermutus *et al*., 2001; Zahnd *et al*., 2004). Unbound or weakly binding library members are then readily eliminated by washing.

The selection of sub-nanomolar affinity binders, however, is usually not additionally favored under such conditions. Therefore, often time-consuming single-clone screening must be carried out. In some cell-based platforms for directed evolution, the screening can be accelerated by the use of cell sorters and fluorescently labeled antigens (Boder *et al*., 2000). However, in iterative affinity maturation of binders with already low nanomolar affinity, more efficient strategies are needed. Because of the size of the molecular repertoire, it is important to apply a selection pressure that is stringent enough to remove undesired library members but not so harsh as to eliminate good clones.

Off-rate selections (Hawkins *et al*., 1992; Jermutus *et al*., 2001; Zahnd *et al*., 2004) have proven to be a powerful tool to select binders with high affinity. Various experimental studies summarized by Northrup and Erickson (1992) (and further analyzed computationally by these authors) suggest that the intrinsic association rate constant for protein–protein interactions in normal salt conditions does not normally exceed $5 \times 10^6\ M^{-1}\ s^{-1}$, making the off rate the key parameter for increasing affinity. The actual panning step is usually performed in solution with biotinylated (or otherwise tagged) antigen. After panning, however, the complexes are not directly isolated, but a large excess of unbiotinylated antigen is added to the panning reaction. The unbiotinylated antigen acts as a competitor and captures the vast majority of the binding molecules that dissociate from the biotinylated antigen after a sufficiently long incubation time. The probability that a binder dissociates from its antigen is mainly defined by its dissociation rate constant. The longer the complexes are incubated with excess competitor, the greater the likelihood that a given complex will dissociate from its biotinylated antigen. The excess amount of competitor prevents rebinding of library members with fast off rates to

biotinylated antigen; thus, capture via biotin enriches for the binders with the slowest off rates in the original pool. The efficiency of this selection principle has been demonstrated several times (e.g. Hawkins *et al*., 1992; Jermutus *et al*., 2001; Zahnd *et al*., 2004). In ribosome display experiments, incubation times of 10 days for peptide-binding single-chain Fv antibody fragments (Zahnd *et al*., 2004) and 25 days for protein-binding DARPins (C.Z. and A.P., unpublished results) have been applied.

While affinity maturation experiments have usually resulted in higher affinity binders, arbitrarily increasing the time of selection for higher stringency is not productive, as we have made the following observations during such experiments: (i) after selection, a surprisingly large fraction of the selected binders did not bind to the cognate antigen at all and (ii) only a minor fraction of the clones that had retained their ability to bind the antigen showed improved binding properties. In addition, it seemed that after a certain duration of incubation, the selection only became worse (E. Wyler and A.P., unpublished results).

The first issue can be corrected with a single panning round without strong selection pressure; this 'binder-collection' round rescues the library. For example, after 10 days of off-rate selection against the peptide GCN4(7P14P), only 7% of the clones still bound to the peptide; after a non-selective panning round, more than 80% bound to the peptide again (Zahnd *et al*., 2004). The very low percentage of library members that still bind the biotinylated antigen after long times may simply be a consequence of the very low numbers of remaining molecules: a nonspecific background signal results from non-binding or non-functional molecules, which are produced during random mutagenesis and are carried through every selection round at a constant rate. Therefore, a non-selective round enriches binding molecules again over this background signal. However, the second issue—the lack of significant affinity maturation during stringent off-rate selection—could not be readily explained or solved.

The behavior of pools of binders under selection conditions is non-trivial to predict. Boder and Wittrup (1998) used deterministic models to estimate optimal equilibrium and kinetic screening conditions for combinatorial cell-surface libraries sorted by fluorescence-activated cell sorting. Here, we build upon their analysis of off-rate selections by examining how ligand rebinding, library complexity and multiple selection rounds can impact enrichment in directed evolution approaches. We first derive an analytical result which identifies the parameters that control enrichment in off-rate selection experiments performed on simple binary libraries when rebinding events are explicitly considered. We then use a Monte Carlo-based approach to elucidate how these parameters impact off-rate selections using more complex libraries. Our results may help to further optimize off-rate selection experiments that are used to enhance molecular affinities.

Methods

Derivation of analytical model result

We refer to the molecule to be recognized as target or antigen, and to the members of the library as binders or ligands. The analytical model assumes that biotinylated antigen (B) is immobilized and in significant excess over the total concentration of library members. After washing to remove all free ligand, unbiotinylated antigen (U) is added in solution (in even greater excess) at the start of the off-rate selection experiment. Holding B and U constant and assuming mass-action kinetics for binding (with the same parameters for biotinylated and unbiotinylated antigen), we can first solve explicitly for $L_1(t)$ (free ligand 1) by integrating the following equation with initial condition $L_1(0) = 0$:

$$\frac{dL_1}{dt} = -k_{on1}L_1(B+U) + k_{off1}(L_{1T} - L_1), \qquad (1)$$

where L_{1T} is the total ligand 1 (free, bound to B, and bound to U). $L_1(t)$ can then be plugged into the following equation to analytically solve for $C_1(t)$ (complex of ligand 1 bound to B):

$$\frac{dC_1}{dt} = k_{on1}L_1B - k_{off1}C_1. \qquad (2)$$

Non-dimensionalization of this result leads directly to Eq. (5) in the main text.

Generation of large library distributions in silico

While the analytical model provides insights into the parameters that most critically impact off-rate selection, a binary library is not representative of a complex initial library nor of those typically generated by random mutagenesis in affinity maturation experiments. Therefore, more appropriate starting libraries were generated *in silico*. Each library consisted of an array: each element in the array represented a library member and was assigned a unique identifier and dissociation rate constant. The dissociation rate constants of the library members were generated randomly, but could be biased by any predefined frequency-distribution. In the simplest type of library, the dissociation rate constant was log-normally distributed, in which case the distribution is defined by two parameters: the average of the negative logarithm of the individual dissociation rate constants $<-\log(k_{off})>_0$ and the variance (σ). $<-\log(k_{off})>_0$ defines how tightly the starting pool already binds to the target on average; the higher $<-\log(k_{off})>_0$ is, the higher the stringency that is needed to enrich for improved binders. The variance is a measure of the diversity of off rates in the pool, and thus quantifies the frequency with which higher affinity binders are present.

It is likely that in many scenarios of experimental affinity maturation, the great majority of binders would have faster off rates upon randomization and generation of the library than the starting clone(s). Therefore, libraries with non-Gaussian off-rate distributions were also generated, specifically bimodal and asymmetric input libraries.

Finally, the size of a given library was adjusted to the length of the simulated selection experiment. After long off-rate selection experiments, only a minor fraction of the initial library was bound to the selectable antigen. To ensure that all selection simulations ended with a statistically significant number of binders, the libraries had to be larger for longer simulations. Our libraries typically contained between 1×10^4 and 2×10^7 binders.

Stochastic simulation algorithm

The dissociation of an antigen-binder complex was assumed to follow first-order kinetics. Consequently, the probability p_{diss} that a complex would dissociate from its antigen in the time interval Δt was defined by the dissociation rate constant k_{off} and could be described by the following equation:

$$p_{\mathrm{diss}} = 1 - e^{-k_{\mathrm{off}}\Delta t} \tag{3}$$

Due to the excess of competitor antigen (and assuming no mass-transport limitations), immediate rebinding of a dissociated ligand molecule to the same antigen can be neglected in determining this probability. However, the probabilistic nature of the rebinding event to either biotinylated or unbiotinylated antigen was indeed simulated (see what follows). Since the concentration of antigen was chosen to be much greater than the K_{D} range of the library and also at least as high as the concentration of complexes, it was assumed that all library members were bound to either biotinylated antigen or unbiotinylated antigen after each time increment. Reducing the concentration of antigen could introduce additional selection pressure on the on-rate of the complexes; however, this effect was not considered here.

Before each simulation, several parameters were defined, including the library distribution (see above), the total duration of the selection experiment (t_{tot}), the ratio of unbiotinylated antigen to biotinylated antigen (U/B) and an integration interval (Δt), which defined the resolution of the simulation.

We used a simple Monte Carlo algorithm, which ran as follows. A dissociation rate constant was assigned to each binder in the library. Every binder was initially set to be bound to biotinylated antigen and its status (bound either to biotinylated or unbiotinylated antigen) was monitored throughout the simulation. For each binder, a random number was generated between 0 and 1, and the probability that this binder would dissociate (p_{diss}) during the integration interval Δt was calculated according to Eq. (3). If the random number was greater than p_{diss}, the ligand would remain in complex to its bound antigen. However, if the random number was smaller, (i) the number of biotinylated antigen molecules increased by one and (ii) the dissociated ligand had to bind to a new antigen. Whether this rebinding event involved a biotinylated or unbiotinylated antigen depended on the relative concentrations of these two species. In our model, the probability of binding to a biotinylated antigen (p_B) is:

$$p_B = \frac{N_B}{N_B + N_U}, \tag{4}$$

where N_B is the number of free biotinylated antigen molecules and N_U is the number of free unbiotinylated antigen molecules. By comparing p_B to a second random number (also between 0 and 1), it could be determined whether the rebinding step would involve biotinylated or unbiotinylated antigen.

The dissociation and binding events within a given integration step were determined for each member in the library. The remaining free biotinylated and unbiotinylated antigen molecules were then enumerated and used in calculating the probabilities in the next integration step. This process was iterated until the simulation time reached t_{tot}.

The algorithm was implemented using the Visual Basic module of Microsoft Excel to enable tracking of each individual library member and to allow direct output of simulation results into spreadsheet format. Off-rate selection simulations were then performed with a predefined library, and the status of the library could be assessed both numerically and graphically at any time during the selection process.

Results

Analytical solution of enrichment ratio of a binary library during off-rate selection

To first gain mechanistic insight into the parameters that most critically impact off-rate selection, we derived an analytical solution for the concentration of a single library member over time. In dimensionless form, the expression is:

$$\theta_1 = e^{-\tau_1} + \frac{\beta_1}{\beta_1 + \mu_1 + 1} \cdot \left[1 - e^{-\tau_1} + (\beta_1 + \mu_1)^{-1}\left(e^{-(\beta_1+\mu_1+1)\tau_1} - e^{-\tau_1}\right)\right] \tag{5}$$

where θ_1 is the fraction of a given library member (ligand 1) that is still bound to biotinylated antigen, τ_1 is dimensional time (t) multiplied by the dissociation rate constant (k_{off1}), β_1 is the concentration of biotinylated antigen (B) divided by the equilibrium dissociation constant of this complex (K_{D1}) and μ_1 is the concentration of unbiotinylated competitor (U) divided by K_{D1}. The non-dimensional independent variable τ_1 can be viewed as a time that is rescaled by the dissociation rate constant for ligand 1. Similarly, the dimensionless parameters β_1 and μ_1 can be thought of as concentrations rescaled to the equilibrium dissociation constant for ligand 1. While the dimensional concentrations B and U are by necessity the same for all binders, the scaled values β and μ describe how strongly B and U actually interact with a given binder because their dimensional concentrations are normalized by the strength of the interaction (K_{D}). Therefore, the enrichment ratio of a binary 'library', comprised of just two unique members, is simply θ_2/θ_1. An important insight from this non-dimensional analysis is that the enrichment ratio only depends on the values of τ, β and μ associated with each library member. Nevertheless, it is more physically meaningful to discuss how the corresponding dimensional quantities (t, B and U) impact the enrichment ratio for simple binary libraries, so we will describe the consequences of varying these dimensional quantities.

Our base model assumes that the two proteins under selection have equal association rate constants ($k_{\mathrm{on}} = 10^5\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$) but unequal dissociation rate constants ($k_{\mathrm{off1}} = 10^{-2}\ \mathrm{s}^{-1}$ and $k_{\mathrm{off2}} = 10^{-3}\ \mathrm{s}^{-1}$); therefore, their K_{D} values are different ($K_{\mathrm{D1}} = 100$ nM and $K_{\mathrm{D2}} = 10$ nM). Of note, the different k_{off} and K_{D} values result in distinct scales for τ, β and μ for the two ligand species.

To understand the effect of competitor concentration on optimal selection conditions, we held the biotinylated antigen concentration constant ($B = 5$ nM) and varied the competitor concentration ($U = 0$–1000 nM) (Fig. 1A). In the absence of competitor ($U = 0$), the enrichment ratio initially increases with time, goes through a maximum ($\theta_2/\theta_1 \sim 12.4$), and then reaches a new equilibrium ($\theta_2/\theta_1 = 7$). The concentration of U has several different effects on the change in

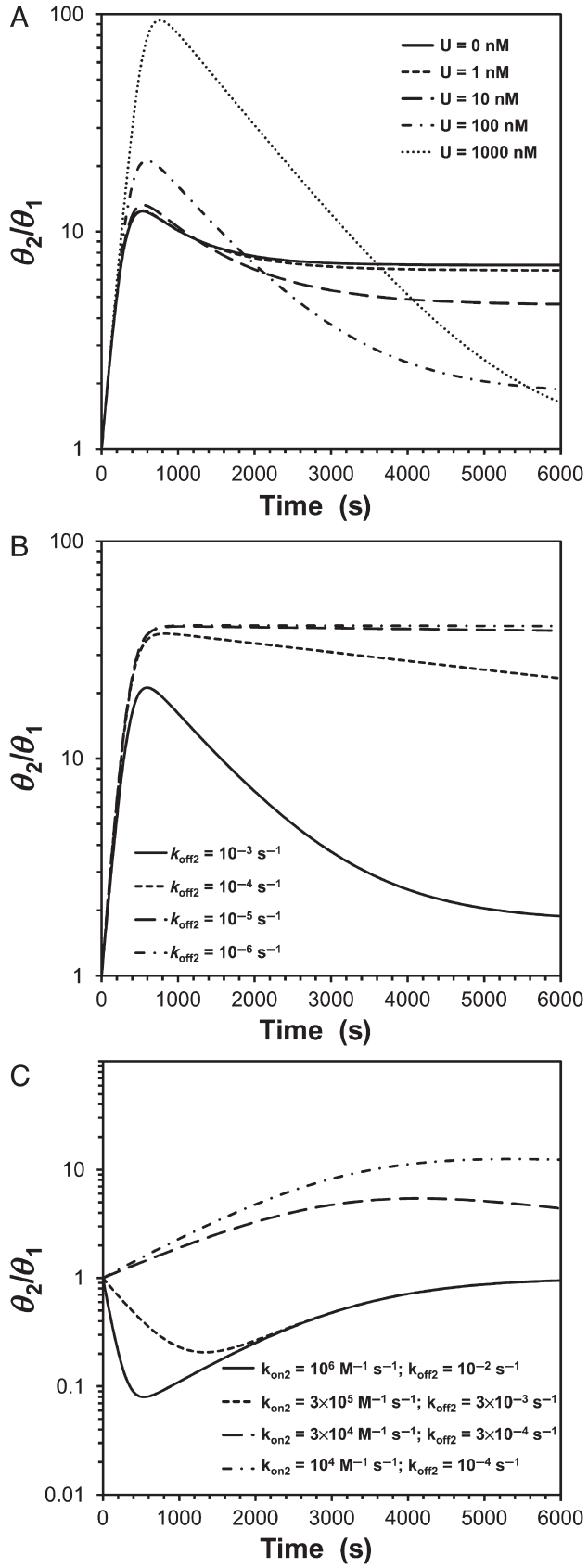

Fig. 1. Analytical simulations of off-rate-based enrichment in binary libraries. Unless otherwise noted, the model parameters are: $k_{off1} = 10^{-2}\ s^{-1}$, $k_{off2} = 10^{-3}\ s^{-1}$, $k_{on} = 10^5\ M^{-1}\ s^{-1}$, $B = 5$ nM, $U = 100$ nM. (**A**) Effect of U (0–1000 nM) on enrichment ratio. (**B**) Effect of k_{off2} on enrichment ratio. (**C**) Effect of binding kinetics on enrichment ratio when k_{on} and k_{off} are varied such that equilibrium affinity is constant.

enrichment ratio over time. At very short times ($t < 200$ s), the change in enrichment ratio is essentially independent of U. This is due to the fact that re-binding of any dissociated molecules to either biotinylated or unbiotinylated antigen is negligible on this time scale. At intermediate times, a higher U not only leads to an increase in the maximum enrichment ratio achievable, but also broadens the time window during which an enrichment ratio greater than 7 (equilibrium θ_2/θ_1 value for $U = 0$) can be achieved. Importantly, if the experiment is carried out for too long and this time window is missed, the enrichment ratio can actually *decrease* with increasing U (e.g. compare θ_2/θ_1 values at $t = 6000$ s in Fig. 1A). This is evident from the equilibrium value of θ_2/θ_1, which can be readily derived using Eq. (5) (taking the limit as $t \to \infty$):

$$\frac{\theta_2}{\theta_1} = \frac{\beta_2(\beta_1 + \mu_1 + 1)}{\beta_1(\beta_2 + \mu_2 + 1)} = \frac{B + U + K_{D1}}{B + U + K_{D2}} \qquad (6)$$

As U (or B) increases significantly above both equilibrium dissociation constants, the enrichment ratio becomes independent of the K_D values and tends towards one. Thus, a very high U is beneficial at intermediate times because it reduces the likelihood of rapidly dissociating ligands rebinding to biotinylated antigen; however, as the reaction approaches equilibrium, this same high U binds all free ligands indiscriminately, thus effectively eliminating the selection pressure.

A second simulation was performed to compare the influence of the dissociation rate constant (k_{off2}) of the higher affinity binder in the binary library (Fig. 1B). As k_{off2} is decreased, the enrichment ratio becomes less sensitive to wait time. Thus, while it is preferable to carry out selections for a time as close to t_{opt} as possible, the margin for error increases as the k_{off} of the desired ligand decreases. However, it should also be noted that all of the curves shown in Fig. 1B do eventually asymptote to a θ_2/θ_1 value of less than 2 as predicted by Eq. (6), so an equilibrium selection would not be fruitful.

Finally, we wanted to simulate whether off-rate selections could be used to discriminate between two ligands with the same K_D value, but with different kinetics of binding (Fig. 1C). In every case, the ligand with slower kinetics (i.e. the smaller k_{off} value) is preferentially enriched, even though the optimal wait time is strongly dependent on the kinetic constants of the ligands. As expected for two ligands with the same K_D, the equilibrium value of θ_2/θ_1 is 1, so discrimination can only be achieved with a kinetic selection.

Sensitivity and specificity of off-rate selections

We further used our analytical model to elucidate how the sensitivity and specificity of off-rate selections change with selection time, ratio of U to B, and background signal (non-specific capture of non-binding or non-functional molecules). In a binary library, in which ligand 1 is unwanted (low affinity) and ligand 2 is desired (high affinity), the following equations describe the true positive probability (TPP) and false positive probability (FPP) during selection:

$$\text{TPP} = \text{sensitivity} = (1 - \text{BG}) \cdot \theta_2 + \text{BG} \qquad (7)$$

$$\text{FPP} = 1 - \text{specificity} = (1 - \text{BG}) \cdot \theta_1 + \text{BG} \qquad (8)$$

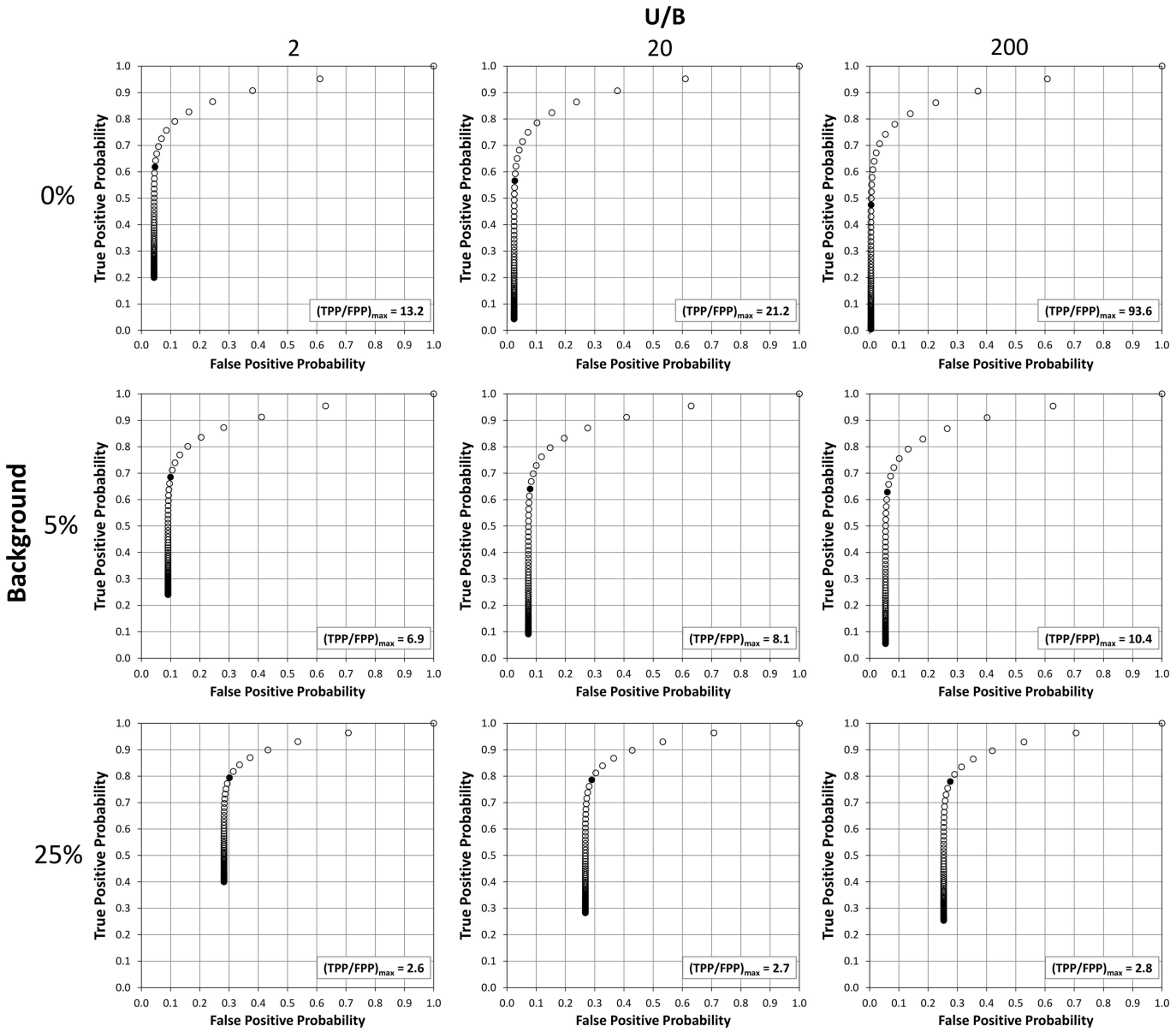

Fig. 2. Receiver operating characteristic (ROC) curves for binary libraries, with parameters as in Fig. 1A. The true positive probability (TPP) (the probability of capturing the desired high-affinity ligand) and false positive probability (FPP) (the probability of capturing the undesired low-affinity ligand) are shown as a function of selection time, i.e. incubation with the competitor. ROC curves were generated for various *U*/*B* ratios (2, 20, and 200) and background signals (0, 5, and 25%). Each off-rate selection starts at the top right corner of the plot (TPP = FPP = 1) and each dot along the trajectory represents a 50 s increment in selection time. In each plot, the selection time corresponding to the maximum enrichment ratio $(TPP/FPP)_{max}$ (whose value is given in the bottom right corner) is shown as a solid dot.

where θ_1 is defined as in Eq. (5), θ_2 is analogously defined, and BG is the fraction of the signal arising from non-specific background. For simplicity, BG is assumed to be constant (and the same value) for both ligands.

Using the same parameters as in Fig. 1A, we plotted the receiver operating characteristic (ROC) curves for different values of *U*/*B* and BG (Fig. 2). The point in the top right corner of each plot represents $t = 0$, the start of the incubation with the competing soluble ligand, where TPP and FPP are both 1 prior to any selection pressure. Following the trajectory of points (each of which represents a 50 s increment of incubation time), we note that each ROC curve is biphasic: there is a brief initial phase during which FPP rapidly decreases (with only a modest decrease in TPP) and thereafter TPP rapidly decreases. The duration of this first phase is relatively independent of *U*/*B* and BG.

Based on the maximum ratio of TPP to FPP in each ROC curve, t_{opt} is shown as a solid point. However, the value of $(TPP/FPP)_{max}$ is substantially different in the various plots: for BG = 0.25 and *U*/*B* = 2, $(TPP/FPP)_{max}$ is 2.6 but for BG = 0 and *U*/*B* = 200, $(TPP/FPP)_{max}$ is 93.6. For comparison, the ROC curves in the top row of Fig. 2 (BG = 0) correspond exactly to three of the curves in Fig. 1A (*U* = 10, 100, and 1000 nM). Examining all of the ROC curves, TPP actually *decreases* as $(TPP/FPP)_{max}$ increases, indicating that maximum enrichment is achieved by minimizing FPP. Notably, increasing *U*/*B* always decreases the lowest attainable FPP for a given BG, thus improving the specificity of

the selection and the enrichment ratio (see also Fig. 1A). In summary, analysis of the ROC curves suggests that, for maximal enrichment of the high-affinity ligand, the highest practical concentration of soluble competitor should be used (highest possible U/B ratio), but the incubation time should not greatly exceed t_{opt}, since the sensitivity of the selection (amount of high-affinity ligand bound) will then decrease significantly without any noticeable gain in specificity (amount of low-affinity ligand bound remains almost constant).

Stochastic simulation of off-rate selection experiments

As highlighted in the analytical model, the kinetic profile of a given library member depends only on its scaled time, τ, and non-dimensional parameters β and μ. In the limit of $\beta \gg 1$ and $\mu \gg 1$ (as was assumed in the stochastic model to safely neglect any free ligand species), this further reduces the dependency to only two entities, τ and the ratio of unbiotinylated antigen to biotinylated antigen ($= \mu/\beta = U/B$), which we examined in greater detail.

Impact of waiting time on selection efficiency

Simulations were first performed with an input library having a log-normal distribution with respect to k_{off} (Fig. 3A). Additional selections were performed using bimodal (Fig. 3B) and asymmetric (Fig. 3C) libraries. As output, the average of the negative logarithm of the individual dissociation rate constants of an enriched pool $<-\log(k_{off})>$ was monitored as a function of time. In addition, the number of library members bound to biotinylated antigen was also monitored since this was expected to be an important metric in assessing selection stringency.

The results suggested that the dissociation behavior of the entire library (Fig. 4A) did not follow a single-order exponential, but more complex decay kinetics. Furthermore, the average of the negative logarithm of the individual dissociation rate constants of the library $<-\log(k_{off})>$ also showed a biphasic response as a function of time (Fig. 4B). Starting from the initial value $<-\log(k_{off})>_0$, this metric increased with time until a maximum value, $<-\log(k_{off})>_{max}$, was reached at time t_{opt}. This value corresponds to the minimal (=optimal) average k_{off}. This result mirrors those from the analytical model and suggests that arbitrarily long waiting times are in fact detrimental to the selection. To understand how the distribution actually changes over time, we followed the library's off-rate distribution over time by halting the simulation at several intermediate time points (Fig. 4C and D).

At relatively short selection times, less than t_{opt}, not only did the average off rate of the enriched library members decrease (and thus $<-\log(k_{off})>$ increase), but the variance decreased (Fig. 4C); both of these effects were expected, since the selection pressure improved the average off rate by reducing the diversity of the pool. Additionally, the distributions were still log-normal. As the waiting times increased beyond t_{opt}, the distributions became bimodal (Fig. 4D). One peak had a $<-\log(k_{off})>$ value that was even higher than before, but a new second peak appeared with a $<-\log(k_{off})>$ that was the same as the input library. As waiting time increased, this second peak became a larger fraction of the total enriched population until it eventually dominated the entire 'selected' population, which was virtually indistinguishable from the input library.

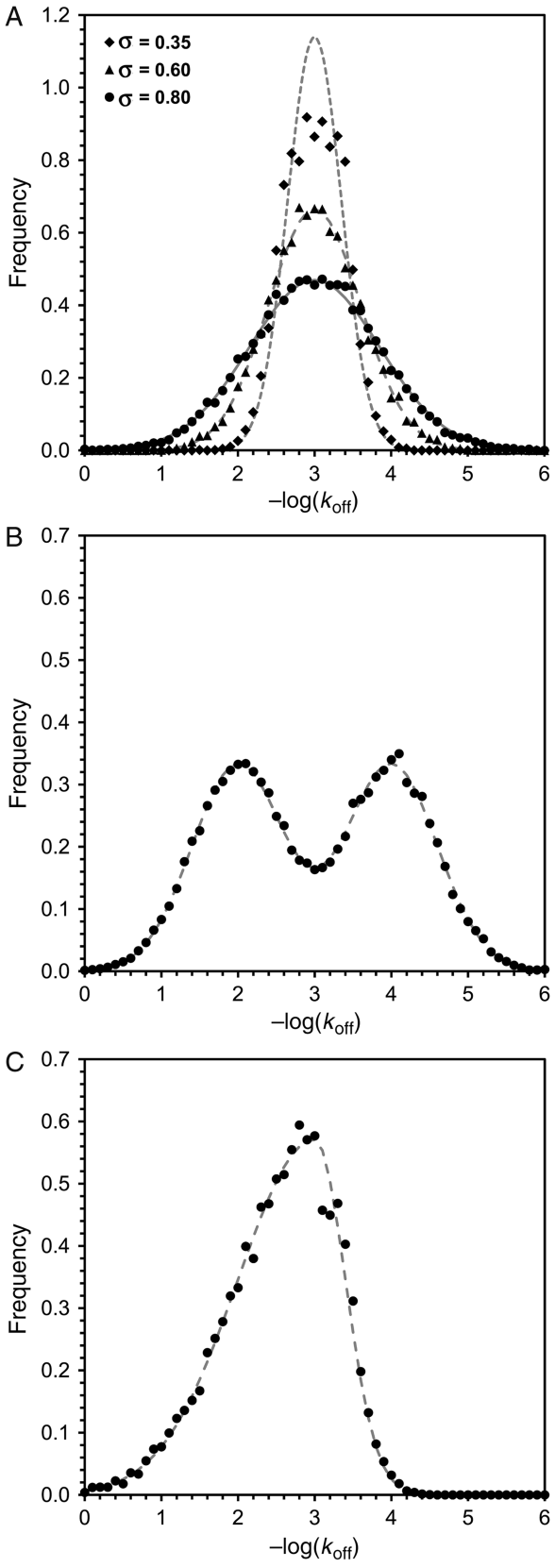

Fig. 3. Frequency distributions of off rates in different libraries. (**A**) Libraries with Gaussian distribution in the logarithm of their off rates with different variances and an $<-\log(k_{off})>_0$ of 3. (**B**) Bimodal library consisting of two superimposed libraries with Gaussian distributed off rates, each having a $\sigma = 0.6$ and peaks at $<-\log(k_{off})>$ of 2 and 4, respectively. (**C**) Library with asymmetric distribution of off rates. All libraries shown contained 65 000 independent binders.

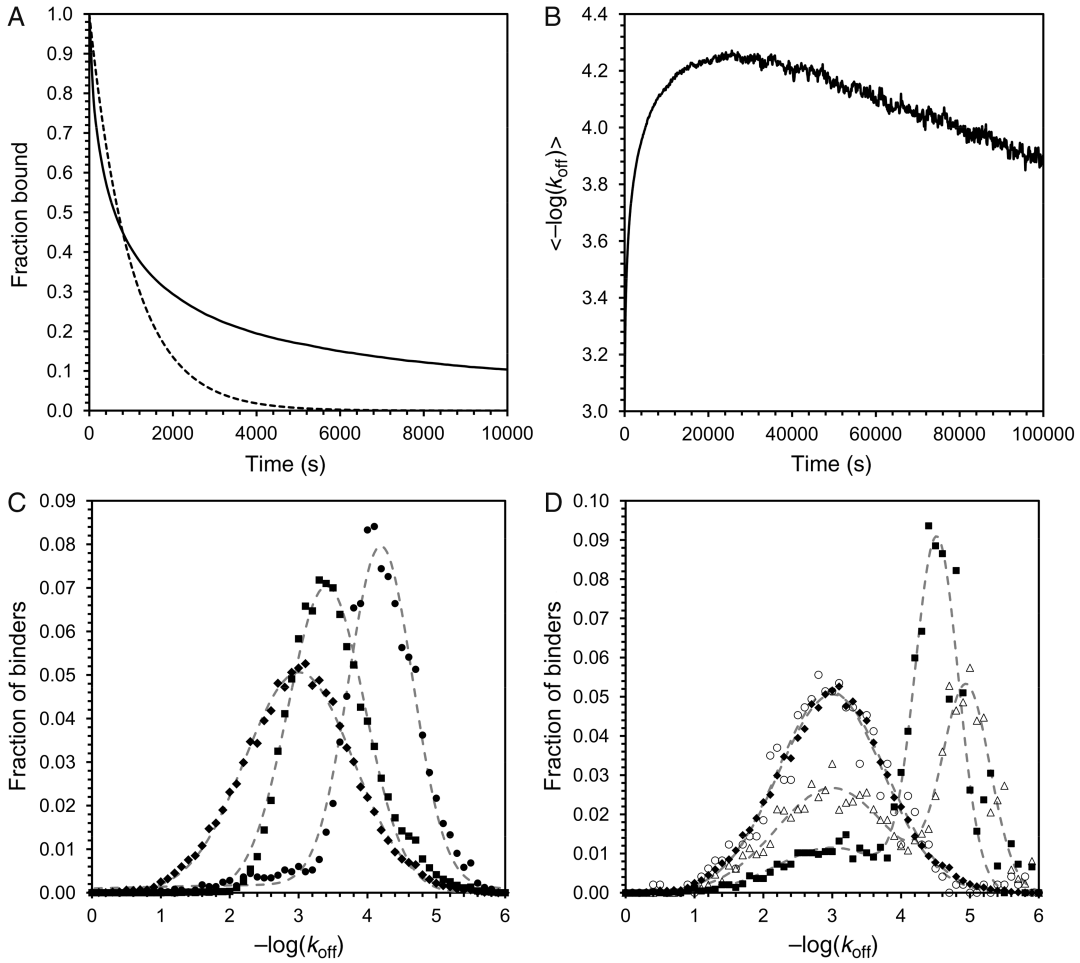

Fig. 4. Simulation of off-rate selection experiments. (**A**) During a selection experiment the binders dissociate from the biotinylated antigen. Only a minor fraction of the pool remains bound to the biotinylated antigen. The dissociation does not follow single-exponential decay and does not converge to zero but to $(1 + U/B)^{-1}$. In panel (A), the fraction of a library, which is still bound to biotinylated antigen, is shown (solid line). The simulation was performed with a library having an $<-\log(k_{off})>_0 = 3$, $\sigma = 0.6$ and $U/B = 10$. For comparison, the single-exponential decay of a binder with an off rate of $10^{-3}\ s^{-1}$ is shown as a dashed line. (**B**) Monitoring the average off rate of a library during selection reveals an early decrease of the average off rate (increase in $<-\log(k_{off})>$). After a certain time, however, $<-\log(k_{off})>$ peaks and again reaches the initial average off rate, because the system is reaching equilibrium. (**C**) Off-rate frequency distributions were monitored in selected pools after different durations of off-rate selection. The peak of the distribution shifted with increasing selection time towards higher average off rates. Closed diamond, input library; closed square, 5000 s; closed circle, 10 000 s. (**D**) Off-rate selections, which are continued for longer times, start to equilibrate. A second peak in the off-rate distribution of the selected libraries appears which perfectly matches the distribution of the input library. This peak increases in magnitude with increasing selection time, whereas the peak of the affinity-matured binders at higher off rates decreases. Closed diamond, input library; closed square, 30 000 sec; open triangle, 100 000 s; open circle, 1 000 000 s (11.2 days).

By comparing these results with the analytical model, it became clear that the selection pressure, which is inherently kinetic in nature, was effectively eliminated by reaching equilibrium at long selection times. This could also be demonstrated by monitoring the dissociation of complexes during the selection experiment. The pool did not continue to dissociate until there were no ligand molecules bound to biotinylated antigen; rather, the fraction of biotinylated antigen-bound library members asymptotically reached a value set by the ratio of biotinylated to unbiotinylated antigen (Eq. (4)). This was further corroborated by the simulation of asymmetric and bimodal libraries (Fig. 3B and C). For these libraries, the distribution of the input library was also restored after very long selection times (data not shown).

Impact of antigen ratio on selection efficiency and waiting time

The effect of the ratio of unbiotinylated antigen to biotinylated antigen (U/B) was assessed by performing several simulations in which all other parameters were held constant. As U/B was increased, a larger $<-\log(k_{off})>_{max}$ was realized (Fig. 5A and B). In Fig. 5A, $<-\log(k_{off})>_0$ was held constant and library diversity, characterized by the variance σ, was varied. As expected, a more diverse starting library results in a higher $<-\log(k_{off})>_{max}$ and, furthermore, a plot of $<-\log(k_{off})>_{max}$ versus $\log(U/B)$ appears linear with a constant y-intercept and a slope that is proportional to σ. In Fig. 5B, $<-\log(k_{off})>_0$ was varied and σ was held constant; a plot of $<-\log(k_{off})>_{max}$ versus $\log(U/B)$ is also linear, but in this case, the slope is roughly constant while the y-intercept shifts with $<-\log(k_{off})>_0$. Combining these observations into a single empirical result, we obtain:

$$<-\log(k_{off})>_{max} = <-\log(k_{off})>_0 + \gamma_1 \sigma \log(U/B) + \lambda_1 \quad (9)$$

where γ_1 and λ_1 are constants fitted to experimental data.

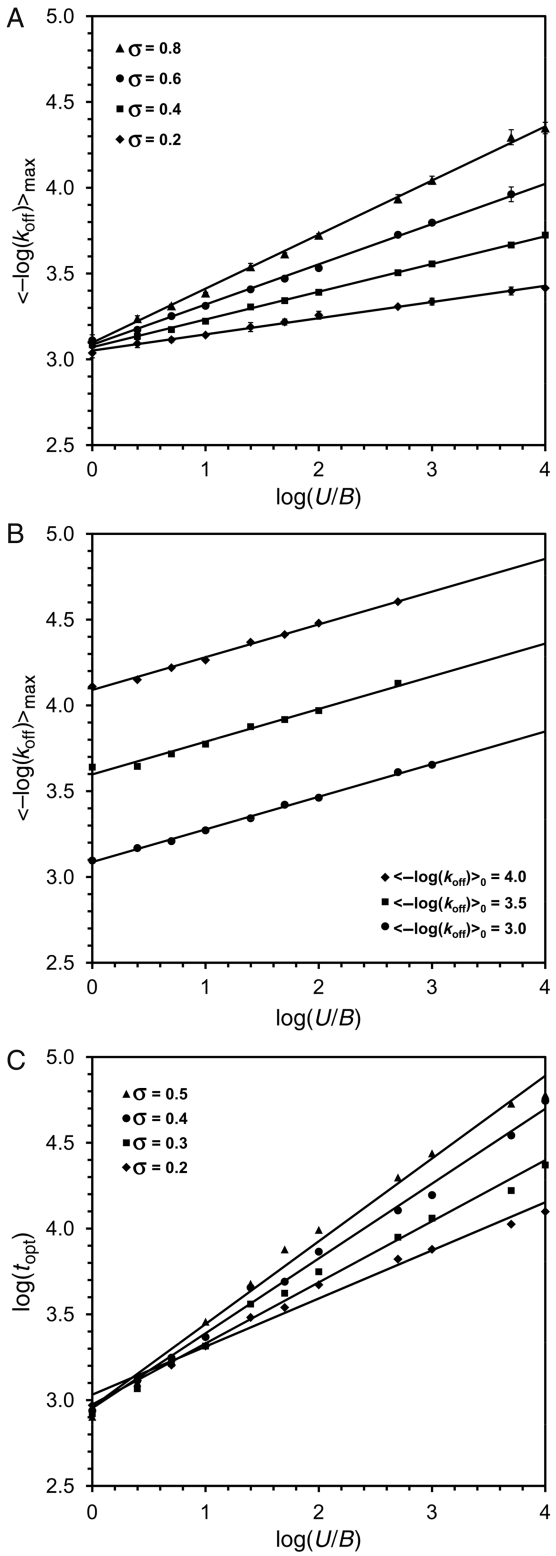

Fig. 5. In-depth analysis of Gaussian-distributed libraries. The effect of different parameters on the outcome of different selection experiments was simulated. All simulations were performed in triplicate and from the simulated curves, t_{opt} and $<-\log(k_{off})>_{max}$ were obtained using a 5-parameter Weibull fit. (**A**) Analysis of the dependency of $<-\log(k_{off})>_{max}$ on U/B with σ as a parameter. Semi-logarithmic plots reveal linear graphs, where the slope is proportional to σ. (**B**) Similarly, the dependency of $<-\log(k_{off})>_{max}$ on U/B with the initial average off rate $<-\log(k_{off})>_0$ as a parameter is shown, influencing the offset of the curves. (**C**) Dependency of t_{opt} on U/B with σ as a parameter. Double-logarithmic plots revealed linear graphs, where again the slopes were dependent on σ and the offsets were dependent on $<-\log(k_{off})>_0$ (data not shown).

We also observed that $\log(t_{opt})$ increased linearly with σ (Fig. 5C), $\log(U/B)$ (Fig. 5C) and $<-\log(k_{off})>_0$ (data not shown). As in Eq. (9), these relationships are best captured by the following empirical equation:

$$\log(t_{opt}) = <-\log(k_{off})>_0 + \gamma_2\sigma\log(U/B) + \lambda_2 \qquad (10)$$

where γ_2 and λ_2 are constants fitted to experimental data.

Thus, for a given input library (characterized by $<-\log(k_{off})>_0$ and σ) and an upper bound on U/B in the experimental setup (typically set by solubility limits or cost constraints), one can use Eqs. (9) and (10) to determine the largest achievable $<-\log(k_{off})>_{max}$ as well as an estimate of the waiting time required. It should be noted that $<-\log(k_{off})>_{max}$ scales linearly with $\log(U/B)$, but t_{opt} scales exponentially with this same quantity.

Rescuing tight binders from overly stringent off-rate selections

The presence of a bimodal distribution of off rates after long simulations (e.g. Fig. 4D) helped to explain some unexpected experimental findings. In very stringent off-rate selection experiments (Zahnd *et al*., 2004; Zahnd *et al*., 2007), only a minor fraction of the analyzed selected clones showed a decreased off rate; most of the clones exhibited off rates that would not have been expected to survive the applied selection pressure. From the ratio of binders with decreased off rates to those with unimproved off rates, further information could potentially be gained. However, this would require an analysis of a very large number of clones and their off rates. In practice, a non-stringent panning round—one in which there is only modest selection pressure, in that the conditions are not stringently optimized (i.e. U/B is not maximized and the selection time is less than t_{opt})—can be subsequently performed to quickly eliminate the subset of the stringently selected library whose off-rate distribution is very similar to that of the starting library. This additional round therefore highly enriches for the affinity-improved species, as revealed by the analysis of single clones.

To see whether the known effect of such a non-stringent panning round could also be captured in our simulations, we mimicked this experimental procedure *in silico*. We first began a simulation of an off-rate selection and then stopped the selection when two major peaks were present in the selected library, one corresponding to the equilibrated input library and one to the affinity-matured library (Fig. 6A). The total number of clones is dramatically decreased at this time point, and this sparser sampling resulted in the coarser distribution profile seen in this figure. This bimodal distribution was 'amplified' (i.e. rescaled) without diversification for use as the input library for a selection round that would be considered non-stringent for the high-affinity distribution but stringent for the low-affinity distribution [since the $<-\log(k_{off})>$ values for these two peaks are substantially different]. This simulation demonstrated that undesired library members from the low-affinity distribution could be effectively eliminated in this step, leading to a significant enrichment of the peak with the desired $<-\log(k_{off})>_{max}$ (Fig. 6A). This result is in good agreement with the experimentally enhanced ratio of affinity-matured clones after a non-stringent selection round (Luginbühl *et al*., 2006; Zahnd *et al*., 2007).

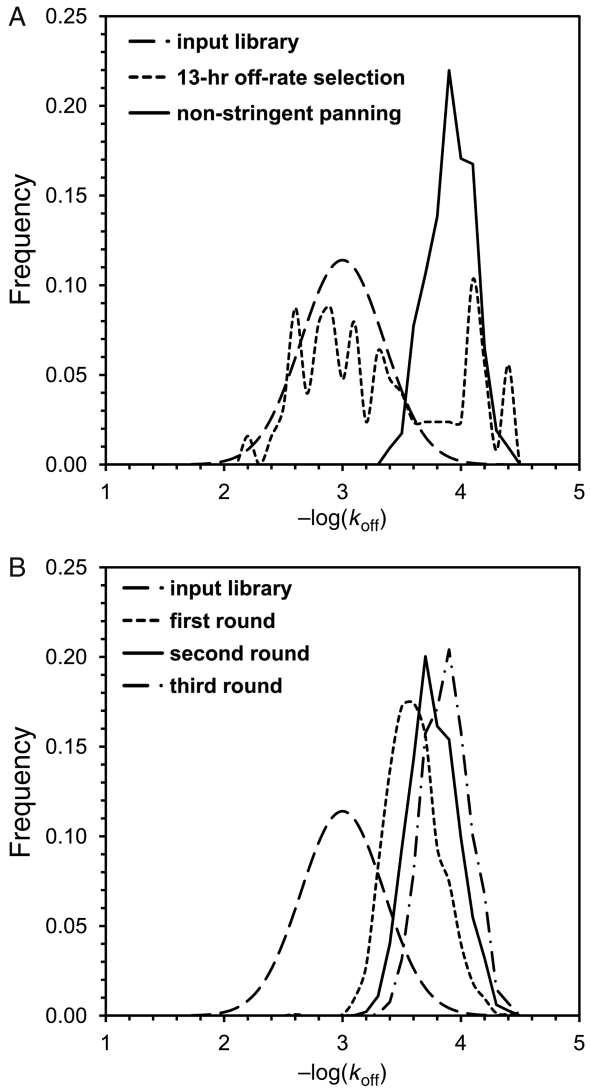

Fig. 6. Comparison of different selection strategies. Two selection strategies were compared side by side starting with the same library, $<-\log(k_{off})>_0 = 3$ and $\sigma = 0.35$. (**A**) A single stringent off-rate selection (13 h) followed by a non-stringent panning round (0.5 h). An equilibration peak appears after the stringent selection, but is effectively eliminated after the non-stringent selection. (**B**) Three consecutive rounds of moderate off-rate selection (2 h per selection). Both strategies result in comparable $<-\log(k_{off})>_{max}$ values for the selected pools. The diversity, however, was >100 times greater using the multi-round selection strategy shown in (**B**) (see main text).

Comparison of selection strategies: tradeoff between stringency and diversity

There has been an ongoing debate about which selection strategy is best for affinity maturation of a given library (Martinez *et al*., 1996; Zaccolo and Gherardi, 1999; Hackel *et al*., 2008). Some have posited that many independent selection rounds with relatively low selection pressure would result in faster enrichment of binders and would allow the sampling of the real diversity of the library. Others have suggested that applying a very rigorous selection pressure on a highly mutated library would not only increase the chance of finding cooperative mutations but also accelerate the selection process.

For a direct comparison of these strategies, the same input library, $<-\log(k_{off})>_0 = 3$ and $\sigma = 0.35$, was used in two independent selection simulations. The first approach was identical to the overly stringent selection strategy described in the previous section: off-rate selection was performed for 13 h with $U/B = 5000$ and then a non-stringent selection round was performed to remove equilibrated binders (Fig. 6A). In parallel, the original library was used to perform a short off-rate selection of 2 h with $U/B = 5000$, and this mild selection procedure was repeated two additional times (Fig. 6B). In both cases, the output library was simply rescaled (without diversification) to generate the input library for the subsequent round. The final $<-\log(k_{off})>_{max}$ was approximately 4 in both simulations, so no significant difference in $<-\log(k_{off})>_{max}$ was observed between the two selection strategies.

We continued the analysis by comparing the diversities of the selected pools. In simulating a very stringent selection round followed by a non-stringent round, an initial library of 5×10^5 binders was used and only 126 ($\sim$0.03%) survived the stringent selection round. In contrast, in the simulation of three successive mild selection rounds, most library members were eliminated in the first round (consistent with the biggest shift towards lower average off rates). Starting with a library of 5×10^4 different binders, 1463 ($\sim$3%) survived the selection in our simulation.

Since a decreased dissociation rate constant is only one of several important parameters in assessing the utility of a clone, the maintenance of greater diversity after selection may enable the selection of sequences that also have other desirable properties, including the epitope targeted, improved association rate constant, thermodynamic stability, expression level or folding efficiency. Thus, iterated but mild selection pressures may provide greater diversity in enriched pools without sacrificing affinity in off-rate selection experiments.

Discussion

The selection of binding molecules with slow dissociation rate constants from complex libraries has become a convenient strategy for both primary selections and affinity maturation of existing binding molecules. In off-rate selection experiments, a pool is incubated with low amounts of biotinylated antigen, before high excess of non-biotinylated antigen is added as competitor. The competitor is generally assumed to prevent rebinding to biotinylated antigen, making the incubation time of the pools in presence of the competitor an important factor determining the selection pressure.

To investigate the parameters that influence efficiency of off-rate selection experiments, we developed both analytical and stochastic models of this process. The analytical model shows that the kinetic profile of a given library member depends solely on three dimensionless quantities—τ, β and μ—each of which can be independently manipulated experimentally by varying the selection time, the concentration of biotinylated antigen, and the concentration of unbiotinylated antigen, respectively. In the limit of saturating antigen (both biotinylated and unbiotinylated), this dependency reduces to two dimensionless quantities, τ and the ratio of unbiotinylated antigen to biotinylated antigen ($=\mu/\beta = U/B$).

We then developed a stochastic framework for understanding the effects of selection time and antigen ratio on the efficiency of mock libraries. These simulations confirmed that overly stringent incubation times actually eliminate the

kinetic selection pressure because the system is at or near equilibrium. The optimal selection time is mainly dependent on U/B and the diversity of the input library. For pools with log-normally distributed k_{off} values, empirical relationships among the selection parameters have been proposed that may be useful in experimental design.

We also addressed, through simulation, whether it would be preferable to perform a single selection round with very high stringency or to perform multiple rounds with a moderate selection pressure. Direct comparison of the two selection strategies revealed no significant difference in the average off rate of the selected pools. However, analysis of the diversity of the selected pools compared with the input libraries revealed that with the stringent selection procedure only a minor fraction of the binders could be recovered than were found after multiple selection rounds with the a less-stringent strategy. Thus, the latter method retains a dramatically greater percentage of the binders with the desired affinity that are actually present in the library and thus allows more choices for subsequent applications.

Due to the artificial character of the simulated libraries and various simplifications, a direct translation of the numerical results to real experiments may be problematic. However, some general insights have been gained, which may be helpful in designing new experiments. First, the largest possible U/B should be used to maximize selection outcome. It should be noted that this can significantly increase the optimal wait time; however, even if the selection is stopped before t_{opt}, the improvement in $<-\log(k_{off})>_{max}$ would be superior to any selection experiment with lower U/B. Second, if the amount of available antigen is not limiting, several subsequent selection rounds with modest selection pressure should be preferable to high-stringency selection rounds because this will lead to higher diversity in the selected pools without sacrificing affinity. Third, if the amount of available antigen is limited and a decreased off rate is the most important property to engineer, a single selection round with the highest possible U/B and longer selection times should be favored to quickly jump to the desired $<-\log(k_{off})>_{max}$. At high U/B, there is a greater margin for error in selection time, which is an important consideration since t_{opt} is generally unknown for complex libraries; however, a single stringent selection will reduce the diversity in the recovered pool, which may adversely affect subsequent protein engineering efforts. Finally, if most of the analyzed clones after random mutagenesis and off-rate selection do not show decreased off rates, the desired ones may be hidden in a large background of unselected or non-functional members. A non-stringent selection round should eliminate the background and enrich these desired clones. In other words, since very few specific clones remain after a stringent selection round, a non-stringent round is frequently necessary simply to amplify these sequences for further manipulations (e.g. subsequent cloning for single-clone analysis).

We conclude with a practical, albeit greatly simplified, summary of our findings. It is not useful to carry out off-rate selections for excessively long times. It would be suitable to start with a selection time that is the reciprocal of the off-rate constant that one is hoping to find in the library (Fig. 1); when in doubt, it is preferable to initially err on the side of shorter selection times (to avoid approaching equilibrium) and then incrementally increase the selection time in later rounds. One should maximize the concentration of competitor relative to immobilized antigen (although the amount of immobilized antigen must remain large enough to allow practical recovery of binders). If the amount of competitor is limiting, a single, stringent selection round, using the highest achievable competitor concentration, should be performed to quickly enrich for binders with the slowest off rate; however, only a handful of different clones may survive this strong selection pressure. If competitor is available in abundance, several selection rounds with a lower selection pressure should be performed to keep the diversity of the enriched library high; this should increase the probability of finding clones with a wider range of biological properties.

Acknowledgements

The authors would like to thank Dr K. Dane Wittrup (MIT) for suggesting the analysis of the data using ROC curves.

Conflict of interest statement. The authors declare no conflict of interest.

Funding

This work was supported by a grant from the University Research Foundation at the University of Pennsylvania to C.A.S. and by a grant from the Schweizerische Nationalfonds to A.P.

References

Binz,H., Amstutz,P. and Plückthun,A. (2005) *Nat. Biotechnol.*, **23**, 1257–1268.
Boder,E. and Wittrup,K. (1997) *Nat. Biotechnol.*, **15**, 553–557.
Boder,E. and Wittrup,K. (1998) *Biotechnol. Prog.*, **14**, 55–62.
Boder,E., Midelfort,K. and Wittrup,K. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10701–10705.
Georgiou,G., Stathopoulos,C., Daugherty,P., Nayak,A., Iverson,B. and Curtiss,R. (1997) *Nat. Biotechnol.*, **15**, 29–34.
Hackel,B., Kapila,A. and Wittrup,K. (2008) *J. Mol. Biol.*, **381**, 1238–1252.
Hanes,J. and Plückthun,A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 4937–4942.
Hanes,J., Jermutus,L., Weber-Bornhauser,S., Bosshard,H. and Plückthun,A. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14130–14135.
Hawkins,R., Russell,S. and Winter,G. (1992) *J. Mol. Biol.*, **226**, 889–896.
Jermutus,L., Honegger,A., Schwesinger,F., Hanes,J. and Plückthun,A. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 75–80.
Luginbühl,B., Kanyo,Z., Jones,R., Fletterick,R., Prusiner,S., Cohen,F., Williamson,R., Burton,D. and Plückthun,A. (2006) *J. Mol. Biol.*, **363**, 75–97.
Martinez,M., Pezo,V., Marlière,P. and Wain-Hobson,S. (1996) *EMBO J.*, **15**, 1203–1210.
Mondon,P., Dubreuil,O., Bouayadi,K. and Kharrat,H. (2008) *Front. Biosci.*, **13**, 1117–1129.
Northrup,S. and Erickson,H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 3338–3342.
Parmley,S. and Smith,G. (1988) *Gene*, **73**, 305–318.
Smith,G. (1985) *Science*, **228**, 1315–1317.
Wong,T., Zhurina,D. and Schwaneberg,U. (2006) *Comb. Chem. High Throughput Screen.*, **9**, 271–288.
Zaccolo,M. and Gherardi,E. (1999) *J. Mol. Biol.*, **285**, 775–783.
Zahnd,C., Spinelli,S., Luginbühl,B., Amstutz,P., Cambillau,C. and Plückthun,A. (2004) *J. Biol. Chem.*, **279**, 18870–18877.
Zahnd,C., Wyler,E., Schwenk,J., Steiner,D., Lawrence,M., McKern,N., Pecorari,F., Ward,C., Joos,T. and Plückthun,A. (2007) *J. Mol. Biol.*, **369**, 1015–1028.